The Pennsylvania State University

The Graduate School

USING ANTS TO FIND COMMUNITIES IN COMPLEX NETWORKS

A Thesis in

Computer Science

by

Mohammad Adi

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science

May 2014

The thesis of Mohammad Adi was reviewed and approved* by the following:

Thang N. Bui
Associate Professor of Computer Science
Chair, Mathematics and Computer Science Programs
Thesis Advisor

*Signatures are on file in the Gradute School.

Abstract

Many systems arising in different fields can be described as complex networks, a collection of nodes and edges. An interesting property of these networks is the presence of communities (or clusters), which represents a subset of nodes within the network such that the connections within these nodes are denser than the connections with the rest of the network. In this thesis, we give an ant-based algorithm for finding communities in complex networks. Ants are used to identify edges which are used to assign the nodes into different clusters. Tests on various synthetic and real-world networks show that the algorithm is able to extract the community structure very well and performs well against other algorithms.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am highly grateful to Dr. Thang N. Bui, my thesis advisor, for his guidance and patience throughout the whole work of this thesis. Also, I would like to thank the member's of the thesis committee, ¡insert names¿, for their valuable feedback.

Chapter 1

Introduction

Complex networks are extensively used to model various real-world systems such as social networks, technological (Internet and World Wide Web) networks, biological networks etc. These networks are modeled as graphs where nodes represent the objects in the system and edges represent the relationship among these objects. For example, in a social network, nodes can represent people and two nodes are connected by a link if they are friends with each other.

These networks exhibit distinctive statistical properties. The first property is the "small world effect", which implies that the average distance distance between vertices in a network is short [4]. The second is that the degree distributions follow a power-law [1], and the third one is network transitivity which is the property that two vertices who are both neighbors of the same third vertex, have an increased probabilibty of being neighbors of one another [6].

Another property which appears to be common to such networks is that of community structure (or clustering). While the concept of a community is not strictly defined in the literature as it can be affected by the application domain, one intuitive notion of a community is that it consists of a subset of nodes from the original graph which between them have a higher density of links as compared to their links with the rest of the graph. In this thesis, we describe an ant-based algorithm for detecting communities in graphs.

Over the course of more than a decade, the task of finding communities in networks has received enormous attention from researchers in different fields such physics,

statistics, computer science etc. As a result, there are currently a vast number of methods which can be used to evaluate the community structure of a network. These methods are described in the next chapter.

Ant algorithms have been used to detect communities in graphs [2] [5] [3]. In our approach, we use artificial ants which traverse the graph based solely on local information and deposit pheromone as they travel. This algorithm uses the cumulative pheromone to build up an initial clustering of the graph. Then a local optimization method is used to reassign the clusters of different nodes based on their degree distribution after which a final step is used to merge clusters and obtain the final partitioning.

The rest of this thesis is organized as follows. Chapter 2 provides more detailed information about the problem statement and covers the previous work done. The ant-based algorithm is described in Chapter 3 and Chapter 4 covers the performance of the algorithm on various synthetic and real-world graphs and compares it to existing algorithms. The conclusion is given in Chapter 5.

Bibliography

[1] Albert-Laszlo Barabasi and Reka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[2] Dongxiao He, Jie Liu, Bo Yang, Yuxiao Huang, Dayou Liu, and Di Jin. An ant-based algorithm with local optimization for community detection in large-scale networks. *CoRR*, abs/1303.4711, 2013.

[3] Di Jin, Dayou Liu, Bo Yang, Carlos Baquero, and Dongxiao He. Ant colony optimization with markov random walk for community detection in graphs. In *Proceedings of the 15th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - Volume Part II*, PAKDD'11, pages 123–134, Berlin, Heidelberg, 2011. Springer-Verlag.

[4] Stanley Milgram. The Small World Problem. *Psychology Today*, 2:60–67, 1967.

[5] S. Sadi, S. Oguducu, and A.S. Uyar. An efficient community detection method using parallel clique-finding ants. In *Evolutionary Computation (CEC), 2010 IEEE Congress on*, pages 1–7, 2010.

[6] D.J. Watts and S.H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, (393):440–442, 1998.